iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67556

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Suicidal Behaviour Detection and Prevention

Sankarana Bhargav[1], Muppidi Jai Bhavani[2], YaminiAppikatla[3], Prof. Dr. G. Krishna Kishore[4]
Department of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology,Andhra Pradesh,India

Abstract: ***The global use of social media has reshaped human relationship with liberty to exchange experience and emotions without limitation. Free communication with the risk of suicidal content, prevalent among the youth generation, requires efficient monitoring and intervention. The project currently utilizes the BERT-based deep learning models to detect and prevent suicide risk based on text analysis. The system assigns the content either as "Sad" or "Not Sad" depending on whether social media postings are being monitored. If the post is categorized to "Sad," the system sends motivational quotes through SMS and Telegram through Twilio and Telethon following an auto-response setting. The system also forwards all distress posts to a suicide help line for immediate expert intervention. Successful Natural Language Processing (NLP) and artificial communication integration ensures successful suicide risk assessment for successful suicide prevention against catastrophic consequences. Suicidal prevention theory of prevention allows for reinforcement of online tracking of mental state and hence is a vital emotional well-being and crisis therapy.***

Index Terms: Social media, natural language processing (NLP), BERT, Twilio, Telethon, suicide detection, suicideprevention.

I. INTRODUCTION

The rise of social networking platforms has transformed the way individuals communicate, offering unprecedented freedom to express emotions and ideas. With such net freedom, however, came more insensitive communications screaming out or even suicide attempts. These helpless victims of self-harm are also reflected in the way they surf the internet and identify themselves and therefore must be vigilant enough so that intervention may be undertaken at the proper time. Prevention of suicide has already been determined ages ago through hand monitoring and self-monitoring and these areWith there now being greater popularity of the social networking websites, individuals now have new ways of communicating with one another with no restraint in expressing emotions and ideas. With such freedom of the net, though, came greater insensitive communications shouting out or even suicide attempts. Such vulnerable victims of self-injury are also exhibited by how they use the internet and identification and hence have to be observant enough so that intervention can be initiated at the right time. Suicide prevention has already been established ages ago with hand monitoring and self-monitoring and these are clumsy and inopportune. In order to overcome such limitations, a project employs artificial intelligence (AI) and natural language processing (NLP) to create an automated suicidal behavior screening and prevention system.

Deep learning technology, while being developed, significantly improved text classification metrics and accuracy rates and enabled AI-driven models to track hundreds of millions of social messages in real-time. Our approach employs a BERT-based approach in sentiment prediction for social media tweets sentiment categorization as "Sad" or "Not Sad." BERT is a transformer-based deep neural network model, which has been shown to be very effective at capturing contextual word-level nuances and thus ideally suited for sentiment classification. In contrast to traditional machine learning where pattern identification must be manually performed, BERT is able to identify patterns in the provided text as part of its process of enhancing class accuracy and transferability. Automated detection of abusive content by our system minimizes cases of human error and is more prevalent and scalable as an anti-suicide initiative.

In addition to reminding the vulnerable users, our care messages also provide live intervention. Our auto-messages remind the user with an auto-care message issued when the post is determined to be "Sad" through SMS and Telegram to reassure him/her that he/she is not isolated and could be helped. Twilio and Telethon APIs run the system so that the messages are sent to individuals on popularly visited web pages. In addition, to ensure early intervention by the mental health practitioners, the system also alerts user information regarding objectionable content to a suicide prevention hotline. In addition to text analysis, with other artificial intelligence and auto dialing features, our solution allows effective suicide interventions and activates early intervention tactics.

This paper presents NLP and deep learning usage on social media suicide detection, particularly the effectiveness of sentiment analysis using BERT in identifying suicidal users. This paper draws solutions to combat the deplorable issue of identifying and intervening in traumatic social media posts to enhance the efficacy rate, accuracy, and scalability of suicide prevention.

As the technology of AI continues to advance, its use in models of mental health will definitely revolutionize suicide prevention by cutting down response time and intervening in high-risk populations with interventions they requireclumsy and untimely. To address such constraints, a project utilizes artificial intelligence (AI) and natural language processing (NLP) to develop an automated suicidal behavior screening and prevention system.

Deep learning technology, during development, vastly enhanced text classification statistics and accuracy rates and facilitated AI-powered models to monitor hundreds of millions of social messages in real-time. Our method uses a BERT-based method in social media tweets sentiment classification as "Sad" or "Not Sad." BERT is a deep neural network transformer-based model that has proven to perform exceptionally well in capturing contextual word-level details and is therefore perfectly suited for sentiment analysis. Unlike conventional machine learning where the pattern recognition needs to be manually executed, BERT can recognize patterns in the input text as part of the class accuracy improvement and transferability in its process. Automated abuse content detection by our system reduces instances of human mistake and is more widespread and reproducible as an anti-suicide program.

Apart from reminding the vulnerable users, our care messages also offer real-time intervention. Our auto-messages remind the user with an auto-care message sent when the post is classified as "Sad" via SMS and Telegram to comfort him/her that he/she is not alone and might be assisted. Twilio and Telethon APIs operate the system so that the messages are sent to people on frequently visited web pages. Besides, in order to trigger early intervention by the mental health professionals, the system also notifies user data about objectionable content to a suicide prevention hotline. Aside from text analysis, with other auto dialing and artificial intelligence features, our solution enables successful suicide interventions and triggers early intervention strategies.

This paper introduces NLP and deep learning application on social media suicidal user detection, specifically the efficiency of sentiment analysis with BERT in suicidal user detection. This paper borrows solutions to counter the despicable phenomenon of detecting and intervening in traumatic social media status to promote the rate of effectiveness, precision, and expansibility of suicidal user detection. With the growing technology of AI, its implementation in mental health models will undoubtedly transform suicide prevention by reducing the time of response and intervening into high-risk individuals with interventions needed.

II. LITERATURE REVIEW

The increasing role of artificial intelligence (AI) and deep learning in mental health detection has led to significant advancements in identifying suicidal ideation through social media posts. Several studies have explored different machine learning and deep learning models for suicide risk assessment, focusing on sentiment analysis, NLP, and multimodal behavioral analysis. These studies highlight the growing importance of AI-driven suicide prevention mechanisms, which can provide timely support and intervention for at-risk individuals.

Ananthakrishnan et al. (2022) proposed the use of multiple BERT-based transformer models, including BERT, DistilBERT, ALBERT, RoBERTa, and DistilRoBERTa, for detecting suicidal intentions in tweets. Their research demonstrated how these pre-trained models can effectively capture contextual language patterns and improve detection accuracy. Similarly, Emily et al. (2024) introduced a hybrid RoBERTa-CNN model that combines RoBERTa's language understanding with CNN's feature extraction capabilities, achieving an impressive 98% accuracy in suicidal ideation detection.

Aldhyani et al. (2022) developed a suicidal ideation detection system using Reddit datasets, incorporating word-embedding approaches such as TF-IDF and Word2Vec alongside hybrid deep learning algorithms. Their study emphasized the importance of using both statistical and contextual methods to analyze social media content accurately. Shaoziong et al. (2020) provided a comprehensive review of machine learning methods for suicidal ideation detection, exploring various NLP techniques and deep learning frameworks. Their research also highlighted ethical concerns and the need for real-world applicability in automated risk assessment models.

Diana et al. (2020) introduced a multimodal approach to suicide risk detection, integrating linguistic, relational, and visual data from social media platforms. Their study demonstrated the effectiveness of combining multiple behavioral indicators for more robust identification. Similarly, Liu et al. (2022) proposed an ensemble deep learning model that incorporated data augmentation techniques to improve the accuracy of suicidal ideation detection. By integrating text-based sentiment analysis with contextual embeddings, their model enhanced classification reliability across various social media datasets.

Ophir et al. (2020) explored the application of artificial neural networks (ANNs) in suicide risk assessment using Facebook posts, analyzing linguistic and psychological cues to detect early signs of distress. Alghowinem et al. (2023) extended this research by focusing on nonverbal behavioral markers through a multimodal region-based behavioral model, which provided additional insights into suicide risk assessment.

Zogan et al. (2020) introduced a hybrid deep learning model, MDHAN, that incorporated textual, visual, and behavioral features to improve explainability in depression detection, an approach that can also be extended to suicidal ideation analysis.

Building upon these advancements, our project employs a BERT-based sentiment classification model to detect suicidal tendencies in social media posts. By leveraging transformer-based NLP techniques, automated sentiment analysis, and real-time intervention strategies, our system aims to provide immediate support to users and alert suicide prevention helplines when necessary. This research contributes to the field of AI-assisted mental health monitoring, offering a scalable and efficient solution for suicide prevention through early detection and timely intervention.

III. MATERIALS AND METHODS

The proposed system utilizes deep learning-based natural language processing (NLP) models to develop an efficient and reliable suicidal behavior detection framework. It leverages a dataset of social media posts, annotated for sentiment classification, to train and evaluate the models. The system implements multiple transformer-based models, including BERT, RoBERTa, to accurately classify posts as "Sad" or "Not Sad." Preprocessing techniques such as tokenization, text cleaning, and stopword removal are applied to enhance model performance. Each model is trained separately, and their outputs are analyzed to determine the most effective model based on accuracy, precision, recall, and F1-score.

Once a post is classified as "Sad," the system triggers an automated intervention mechanism. Supportive messages are sent to users via Twilio (SMS) and Telethon (Telegram) to offer immediate emotional support. Additionally, the details of users posting distressing content are forwarded to a suicide prevention helpline to ensure timely professional intervention. The real-time nature of this system makes it a valuable tool for mental health monitoring, helping to bridge the gap between online distress signals and real-world support.

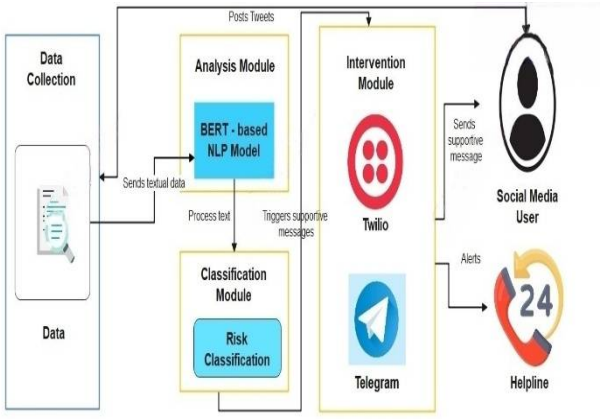


Fig.1 Proposed Architecture

The system architecture (fig.1)for suicidal behavior detection and intervention comprises three key stages: data collection, sentiment analysis, and automated intervention. Social media posts are extracted, preprocessed, and tokenized for analysis. Multiple transformer-based models, including BERT, RoBERTa, classify posts as "Sad" or "Not Sad," ensuring high accuracy. If a post indicates distress, the system triggers real-time support through Twilio (SMS messaging), Telegram (chatbots), and alerts to mental health helplines. This AI-driven approach enhances scalability, accuracy, and automation, enabling proactive suicide prevention by bridging online distress signals with timely professional intervention.

1) Dataset:

The system leverages a dataset of social media posts labeled with emotional sentiments to train and evaluate BERT-based NLP models for detecting distress and suicidal ideation. The dataset undergoes preprocessing techniques such as tokenization, stop-word removal, and vectorization to enhance text representation. By incorporating diverse linguistic expressions of distress, the dataset improves classification accuracy, ensuring early and reliable detection of at-risk individuals. This structured approach enables the development of an effective intervention mechanism, facilitating real-time communication with users and alerting mental health professionals for timely support.

2) Text Processing:

Text preprocessing is performed by tokenizing and embedding social media posts using transformer-based models. Sentiment classification is conducted using multiple models, including BERT, RoBERTa, ensuring robust performance. The text data is converted into numerical representations using deep learning tokenizers, allowing for precise contextual analysis. Further processing involves applying confidence scores and threshold tuning to reduce false positives, improving the reliability of distress detection.

3) Data Augmentation:

To improve model generalization, data augmentation techniques such as synonym replacement, back translation, and sentence shuffling are applied to the text dataset. Synonym replacement substitutes words with similar meanings, enhancing the model's ability to recognize varied expressions of distress. Back translation introduces linguistic diversity by translating text into another language and back, altering phrasing while preserving meaning. Sentence shuffling rearranges words or phrases to prevent overfitting to fixed patterns. These techniques mitigate biases, improve adaptability to different writing styles, and ensure the model accurately detects suicidal tendencies across diverse user-generated content, enhancing the effectiveness of automated intervention efforts.

4) Algorithms:

RoBERTa(Robustly Optimized BERT Pretraining Approach): RoBERTa (Robustly Optimized BERT Pretraining Approach) is a deep learning model designed to enhance performance by training on larger datasets with more dynamic masking techniques. This allows it to better understand nuanced language patterns in social media posts, making it highly effective for detecting suicidal ideation. By leveraging an improved training strategy, RoBERTa achieves higher precision and recall, ensuring more accurate classification of distress signals. Its ability to capture subtle variations in text enables it to minimize false positives and false negatives, making it a reliable tool for identifying at-risk individuals and supporting timely mental health interventions.

IV. EXPERIMENTAL RESULTS

1) *Precision:* Precision measures the fraction of correctly classified "Sad" posts among all posts predicted as "Sad." A high precision value indicates that the model minimizes false positives, ensuring that only genuinely distressing posts are flagged for intervention. It is calculated as:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$$

2) *Recall:* Recall is a metric in machine learning that measures the model's ability to correctly identify all "Sad" posts. It measures how well the system captures true distress signals. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

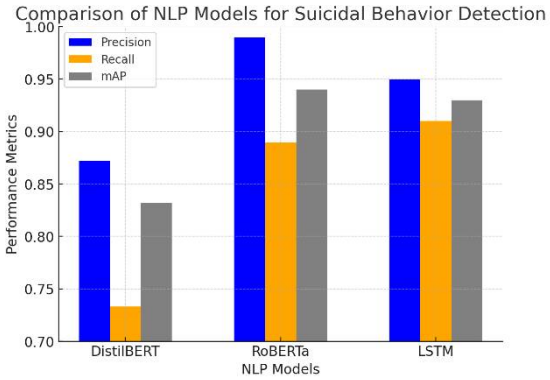
$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

3) *MAP:* Mean Average Precision (MAP) is used to assess the ranking quality of detected distress signals. It considers both precision and recall across different confidence thresholds. MAP at K is calculated as an arithmetic mean of the Average Precision (AP) at K across all users or queries.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

MODEL	Precision	Recall	mAP
DistilBERT	0.872	0.733	0.832
RoBERTa	0.99	0.89	0.94
LSTM	0.95	0.91	0.93

Table.1compares the NLP models using precision, recall, and mAP, highlighting RoBERTa achieves the highest mAP and best precision.



Graph.1 visualizes precision in blue, recall in orange, and mAP in grey for the NLP models, showing RoBERTa achieves the highest mAP and best precision.

tweetid	username	tweet	phone number
3208437	@mike409	Work is going great, productive day.	9017928094
4460483	@jordan189	Loving this new book I'm reading.	9577857967
5107422	@taylor895	I am so sick and tired of living.	9400339825
6789543	@chris610	Excited for the weekend plans!	9047980565
7938145	@morgan773	Feeling blessed and happy today.	9289364282

Fig.2Input tweets

tweetid	username	tweet	phone number	Sentiment
3208437	@mike409	Work is going great, productive day.	9017928094	Not Sad
4460483	@jordan189	Loving this new book I'm reading.	9577857967	Not Sad
5107422	@taylor895	I am so sick and tired of living.	9400339825	Sad
6789543	@chris610	Excited for the weekend plans!	9047980565	Not Sad
7938145	@morgan773	Feeling blessed and happy today.	9289364282	Not Sad

Fig.3Classified tweets

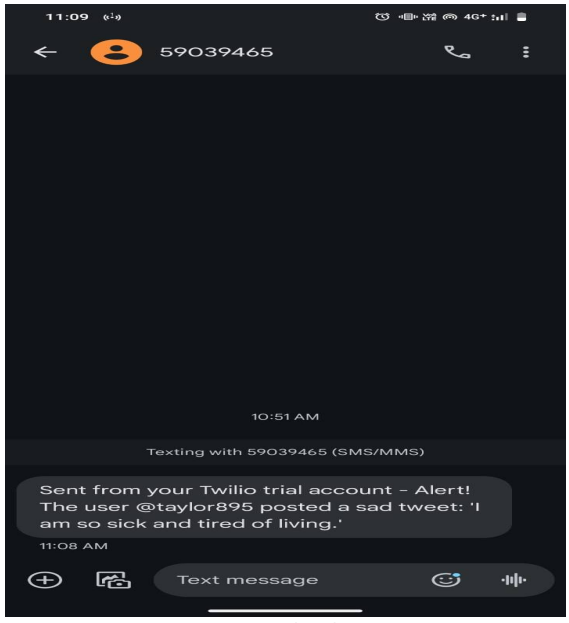


Fig.5 User SMS sent

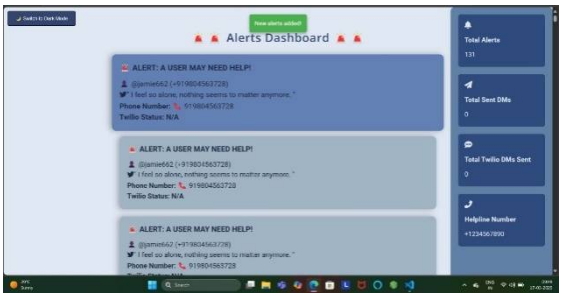


Fig.6 Helpline

V. CONCLUSION

Conclusion

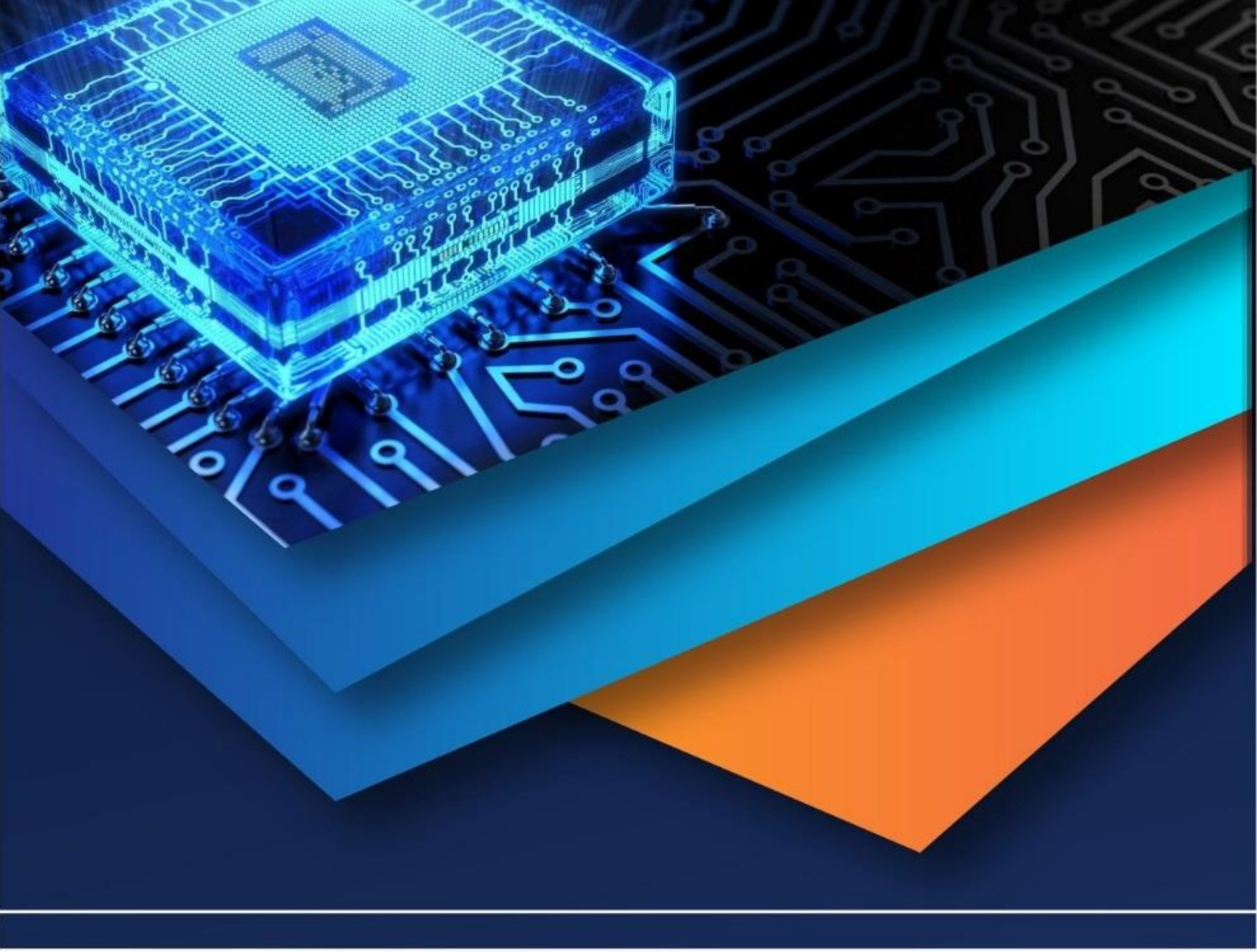
Suicidal behavior detection plays a crucial role in mental health intervention, where timely identification can significantly improve support and prevention efforts. Deep learning-based Natural Language Processing (NLP) models, particularly transformer-based architectures, have demonstrated their effectiveness in analyzing social media posts for distress signals. This study explores the application ofRoBERTa, to develop an efficient suicidal behavior detection system. A curated dataset of social media posts labeled with emotional sentiments is used to train and evaluate these models, ensuring robust learning and accurate predictions. The dataset undergoes preprocessing techniques such as tokenization, stop-word removal, and augmentation methods like synonym replacement and back translation to enhance detection performance. Each NLP model is assessed for its efficiency, classification accuracy, and reliability in detecting suicidal tendencies. The results indicate that advanced transformer architectures significantly improve detection accuracy, making them suitable for real-time mental health monitoring. Performance metrics such as precision, recall, and mean Average Precision (mAP) were used for evaluation, with RoBERTa achieving the highest precision. By automating the detection process, this system reduces dependency on manual assessment, minimizing oversight and assisting mental health professionals in early intervention. The study highlights the potential of deep learning in revolutionizing mental health support, offering a scalable and effective solution for suicide prevention.

*Futurework*includes multimodal emotion detection, real-time monitoring across platforms, AI-driven personalized responses, and integration with mental health services for real-time intervention. Enhancing data privacy through encryption and secure APIs ensures compliance with global laws, strengthening the system's reliability and ethical AI-driven mental health support.

REFERENCES

[1] Parsapoor, M., Koudys, J. W., & Ruocco, A. C. (2023). Suicide risk detection using artificial intelligence: The promise of creating a benchmark dataset for research on the detection of suicide risk. Frontiers in Psychiatry, 14, 1186569.
[2] Tang, H., Rekavandi, A. M., Rooprai, D., Dwivedi, G., Sanfilippo, F., Boussaid, F., &Bennamoun, M. (2023). Analysis and evaluation of explainable artificial intelligence on suicide risk assessment. Scientific Reports, 13(1), 1-12.
[3] Ormerod, C. M., Patel, M., & Wang, H. (2023). Using language models to detect alarming student responses. arXiv preprint arXiv:2305.07709.
[4] Elsayed, N., ElSayed, Z., & Ozer, M. (2024). CautionSuicide: A deep learning-based approach for detecting suicidal ideation in real-time chatbot conversations. arXiv preprint arXiv:2401.01023.
[5] Badian, Y., Ophir, Y., Tikochinski, R., Calderon, N., Klomek, A. B., & Reichart, R. (2023). A picture may be worth a thousand lives: An interpretable artificial intelligence strategy for predictions of suicide risk from social media images. arXiv preprint arXiv:2302.09488.
[6] Chen, Y., Li, J., Song, C., Zhao, Q., Tong, Y., & Fu, G. (2024). Deep Learning and Large Language Models for Audio and Text Analysis in Predicting Suicidal Acts in Chinese Psychological Support Hotlines. arXiv preprint arXiv:2409.06164.
[7] Ananya G., Bhumika J., Minakhi R.(2023). Prediction of Suicide Rates in India Using Machine Learning. 2023 IEEE International Conference on Advances in Science and Technology (ICAST), 1-6.
[8] Maini, M., Srivastava, P., Soni, H., Hemprasanna, & Pillai, A. S.(2023). Prediction of Suicide Rates in India Using Machine Learning. Proceedings of the 2023 IEEE International Conference on Advances in Science and Technology (ICAST), 1-6.
[9] Wibhowo, C., & Sanjaya, R. (2021). Virtual Assistant to Suicide Prevention in Individuals with Borderline Personality Disorder. Proceedings of the 2021 International Conference on Control, Automation, and Information Sciences (ICCAIS).
[10] Qiu, H., Ma, L., & Lan, Z. (2024). PsyGUARD: An automated system for suicide detection and risk assessment in psychological counseling. arXiv preprint arXiv:2409.20243.
[11] Tank, C., Mehta, S., Pol, S., Katoch, V., Anand, A., Jaiswal, R., & Shah, R. R. (2024). Su-RoBERTa: A semi-supervised approach to predicting suicide risk through social media using base language models. arXiv preprint arXiv:2412.01353.
[12] Liu, S., Lu, C., Alghowinem, S., Gotoh, L., Breazeal, C., & Park, H. W. (2022). Explainable AI for suicide risk assessment using eye activities and head gestures. arXiv preprint arXiv:2206.07522.

[13] Liu, M., & Liu, S. (2021). Stability Analysis of a Stochastic SIRS Epidemic Model with a General Nonlinear Incidence Rate. Proceedings of the 2021 IEEE International Conference on ArtificialIntelligence and Computer Applications (ICAICA), 945-949.
[14] Cui, Z., Lei, C., Wu, W., Duan, Y., Qu, D., Wu, J., Chen, R., & Zhang, C. (2024). Spontaneous speech-based suicide risk detection using Whisper and large language models. arXiv preprint arXiv:2406.03882.
[15] Nfissi, A., & Ouni, A. (2024). Unlocking the emotional states of high-risk suicide callers through speech analysis. Proceedings of the 18th IEEE International Conference on Semantic Computing (ICSC).
[16] Nock, M. K., Borges, G., Bromet, E. J., Cha, C. B., Kessler, R. C., & Lee, S. (2008). Suicide and suicidal behavior. Epidemiologic Reviews, 30(1), 133–154.
[17] Atmakuru, A., Shahini, A., Chakraborty, S., Seoni, S., Salvi, M., Hafeez-Baig, A., Rashid, S., Tan, R. S., Barua, P. D., Molinari, F., & Acharya, U. R. (2024). Artificial intelligence-based suicide prevention and prediction: A systematic review (2019–2023). Biomedical Signal Processing and Control, 85, 104965.
[18] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems, 7(3), 1145-1158.
[19] Singh, P. K., & Kumar, A. (2023). Designing a model for suicidal behaviour detection using machine learning. International Journal of Advanced Research in Computer Science, 14(3), 123-130.
[20] Levkovich, I., & Omar, M. (2024). Evaluating BERT-based and large language models for suicide detection, prevention, and risk assessment: A systematic review. Journal of Medical Systems, 48(1), 113.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY